N89-26599

# THE UTILIZATION OF NEURAL NETS IN POPULATING AN OBJECT-ORIENTED DATABASE

William J. Campbell
National Space Science Data Center
Goddard Space Flight Center
Greenbelt, Maryland 20771

Scott E. Hill and Robert F. Cromp
Science Applications Research Inc.
National Space Science Data Center
Goddard Space Flight Center
Greenbelt, Maryland 20771

## ABSTRACT

Existing NASA supported scientific data bases are usually developed and managed by a team of database administrators whose main concern is the efficiency of the data bases in terms of normalization and data search constructs. The populating of the data base is usually done in a manual fashion by row and column as the data becomes available, and the data dictionary is usually defined by the same team (at times with little input from the end science user). This process is tedious, error prone and self-limiting in terms of what can be described in a relational Data Base Management System (DBMS). The next generation Earth remote sensing platforms (i.e., Earth Observation System, EOS), will be capable of generating data at a rate of over 300 Mbs per second from a suite of instruments designed for different applications. What is needed is an innovative approach that creates object-oriented databases that segment, characterize, catalog and are manageable in a domain-specific context and whose contents are available interactively and in near-real-time to the user community. This paper describes work in progress that utilizes an artificial neural net approach to characterize satellite imagery of undefined objects into high-level data objects. The characterized data is then dynamically allocated to an object-oriented data base where it can be reviewed and accessed by a user. The definition, development, and evolution of the overall data system model are steps in the creation of an application-driven knowledge-based scientific information system.

## Introduction

One of the most significant technical issues that NASA must address and resolve is the problem of managing the enormous amounts of scientific and engineering data that will be generated by the next generation of remote sensing systems, such as the Hubble Space Telescope (HST) and the Earth Observation System (EOS). The amount of data these sensors are expected to produce will be orders of magnitude greater than NASA has ever experienced. Consequently new solutions must be developed for managing, accessing and automatically inputting the data into a database in some expressive fashion that will provide a meaningful understanding and effective utilization of this data in a multidisciplinary environment.

Presently, scientific data provided by satellites and other sources (i.e., in situ measurements) are processed, cataloged, and archived according to narrow-mission or project-specific requirements with little regard to the semantics of the overall research. Scientists therefore lack knowledge of or access to potentially valuable data outside of their own field and usually access to this data is long after the actual generation of that data. What is needed is a methodology that will extract and characterize a processed data stream from a remote sensing instrument, and automatically augment appropriate data catalogs for remote browsing, at a high level of abstraction, that would be of interest to NASA's scientific community.

## Concept

The concept is for the system to intercept a data stream from a remote sensing instrument and pass the data through a series of artificial neural networks that have been knowledge engineered or tuned to specifically identify and characterize data objects at a high level of abstraction using the appropriate domain-specific program. These networks or characterization agents will be controlled by a knowledge-based planner and controller that directs the identification and abstraction of objects determined to be of interest to the scientific community. These networks will run in parallel and will be activated as appropriate (predetermined in the tuning process) for the given defined context, for example, for a specific instrument or data stream. Initial passes of these networks characterize data objects at a high level, determined by a threshold level of confidence of a given object to be that object. The information obtained would automatically be inserted into an object-oriented database which builds, indexes, and maintains sets of elements and allows a user to retrieve these data as individual items or as any aggregate of objects. Once a remotely sensed data set has undergone some level of preprocessing (e.g., decompression, radiance values generated, etc.) then additional ephemeris information such as date, time and sensor ID can be added to the parent object or any subdivision of objects.

This data, along with associated meta data and data set identification will then be sent to an appropriate archive. A reference data frame that is specific to a particular domain (science or sensor specific) will then created within the context of the domain world model of the information management system. An important point concerning this process is that much of the information required to catalog and characterize the data set will already be in the knowledge base as a consequence of a priori knowledge acquisition from both the ephemeris information specific to a given sensor as well as the science information a particular sensor is designed to capture [CAM88].

Using the above approach, raw science and engineering data can be efficiently processed and stored using meaningful representations that are more suitable to a user's reasoning. The definition, development, and evolution of the meta frames, agents and overall data system model are the first steps in the evolution of an application-driven knowledge base.

## Design Considerations

The design of a data cataloging and characterization system is predicated on having preexisting knowledge of the domain, the sensor devices and the interpretation of their measurements. It is fruitless to identify, store, and manipulate data if there are no guidelines that differentiate between good and bad observations, or if the integrity of the database cannot be guaranteed.

The suggested design melds subsymbolic processing by a neural network with high-level symbolic processing controlled by an expert system. This design requires a research and development effort in the following areas:

1. Architecture of a neural network which can characterize the pixels in a remotely sensed image based upon the satellite's primary bands.

2. Effective training procedures of a neural net to maximize its performance while minimizing the amount of required computer CPU time.

3. Combination of the technologies of neural network computing and expert systems.

4. Categorization of large data sets in near real time by using an associative memory model as defined by a neural network.

5. Use of an expert system that uses contextual information, such as time of year and location of image, to judge and refine the output of the neural network.

6. Use of an expert system to instantiate an object so that its representation is suitable for a database. This requires a mapping of the characterization of the image data (represented by a subsymbolic collection of pixels) to an object (represented by a symbolic collection of attribute-value pairings).

## Approach

Initial research into steps 1 and 2 of the just outlined research plan will now be presented. First, we will introduce back-propagation, the type of neural network believed best suited to this task. A methodology for and the results of several experiments will be described. The conclusion will be drawn from those experiments that this style of computation appears quite favorable for the categorization of LANDSAT-4 images.

In the past few years, a style of computation termed neural networks has become popular. Perhaps the most successful type of neural network has been back-propagation [RUM86], a supervised learning procedure for training layered networks of neuron-like nodes. A layer of nodes are those nodes which are similarly connected to other layers in a network. Back-propagation networks have an input layer, an output layer, and from one to many intermediate, hidden layers. Connections are unidirectional links between two nodes (called the "from" and "to" nodes); each connection has an associated value called "weight." Each node has an activation value which is a function of the activation values of the nodes connected to it and the weights associated with those connections . There is some flexibility in the form of that activation function, the one chosen for this study is

$$o_j = \frac{1.0}{1.0 + e^{-\left(\sum_i w_{ij}\, o_i + \theta_j\right)}} \tag{1}$$

where

$w_{ij}$ is the weight of the connection linking the $i^{th}$ node to the $j^{th}$ node,

$\theta_j$ is the threshold value for the jth node, and

$o_i$, $o_j$ are the activation values of the $i^{th}$, $j^{th}$ nodes,

and the summation is over all the "from" nodes i connected to the $j^{th}$ node.

The activation value of the input nodes is set by the user as the desired input pattern. That activity then propagates forward, layer by layer, through the network as dictated by the network connectivity. The threshold value for a node acts as a weight from a node with a constant activation value of unity.

During learning, weights are adjusted so as to minimize a measure of the difference between the actual values of the output nodes (the output vector) and the desired values of the output nodes (the target vector) when the network is presented with the input vector to the input nodes and that activity is propagated forward. To do so requires a training set of input vectors and their associated target vectors. Training of the network proceeds in a series of two stage events: first, an input vector is presented to the input nodes and activation is fed forward through the network to produce an output vector. This output vector is compared with the desired target vector and the difference between the two vectors, the error vector, is computed. The measure of error used in this study is

$$E_p = \frac{1}{2}\sum_j (o_{pj} - t_{pj})^2 \tag{2}$$

where

$E_p$ is the error after the $p^{th}$ training pattern,

$t_{pj}$ is the target activation value for the jth node in the output layer in the pth training pattern, and

$o_{pj}$ is the activation value of the $j^{th}$ node after presenting the input vector in the $p^{th}$ training pattern and propagating activity forward.

Given these equations of activity propagation and error measurement, the derivative of the error for a unit with respect to the weights connected to that unit can be recursively calculated. Those derivatives are used to change the weights of the connections between the nodes in the network so as to reduce E upon subsequent presentation of the input vector. By repeating this forward propagation of activity and backward propagation of weight changes for each member of the training set, the connection weights slowly change to a configuration that, upon presentation of each member of the training set of vectors, produce in the output nodes, the target vector which is the correct interpretation for the current input vector. Training of these networks can require a long time, with many presentations of each member of the training set. However, once a neural network is trained, the weights can be implemented in real time software/hardware systems [FOG88]. Complex preprocessing stages can negate this advantage of the neural network approach, however.

The network trains in a series of epochs. Each epoch consisted of the presentation of one input vector in the training set from each category. Subsequent epochs use successive pixels from each category. In this way, if there are 85 pixels with type 2 in the training set, then each of those 85 pixels will be presented once to the network every 85 epochs. An alternative would be to sequentially present each pixel in the training set irrespective of category. The latter method would train pixels in the same ratio as their occurrence in the training set. Although this might be favorable to the former method in terms of the overall percentage correctly classified, it does so at the expense of less well represented categories. The method adopted trains on category type 4 as often as on category type 9.

Both damping and momentum factors are used. The damping factor premultiplies the specified weight change for a connection. The damping factor is the product of a network damping factor, 0.5 in these experiments, and the inverse of the fanin of the "to" unit for the current connection. The fanin of a unit is the number of connections which converge to the unit. The damping factor for a connection is multiplied by the weight change for that connection indicated by the current pixel presentation. The momentum factor premultiplies the accumulated weight change indicated by the previous epoch before accruing the weight changes indicated by the current epoch. The effect of the momentum factor is to smooth out the change in weights between presentations of training pixels.

In this research, values for the first four spectral bands from a LANDSAT-4 Thematic Mapper image are used as input to a back-propagation neural network. The bandpass for those spectral bands are 0.45-0.52 υm, 0.52-0.60 υm, 0.63-0.69 υm, and 0.76-0.90 υm, respectively. Each picture element (pixel) in the image is representative of a 30 x 30 meter area and quantized to 256 levels. The LANDSAT-4 imagery of the region was obtained in July, 1982 [TIL89]. The network is trained to associate the spectral data of each pixel with one of seventeen possible land cover or land use categories.

The network is trained to associate the spectral data from a pixel with the land cover or land use category for that pixel. There are a total of 21,273 pixels of valid ground truth provided; each is encoded as a one byte integer ranging from 1 to 17 (see Table 1). These pixels are contained within a 151 x 151 pixel region within the area designated as subregion 1 in a study by Williams, et al. [WIL84]. This area is about 25 miles SSE of Washington D. C. The land use and land cover data (ground truths) for the region were obtained by photo interpretation of color infrared aerial photography (1:40,000) that was collected over the area on July 13, 1982 and verified by subsequent field visits in October, 1982 [WIL84]. A 15 meter minimum mapping unit criterion was used. However, as that study states, "in the case of agricultural fields, the minimum mapping unit was utilized only to separate one field from another; no attempt was made to delineate within-field variability." Land cover categories were substituted for land use categories in "situations where the land cover components of the categories (e.g. the roofs, lawns, trees, and concrete/asphalt areas of a residential neighborhood) occupied areas with spatial dimensions approximately equal to or smaller than the 15-m minimum mapping unit." Notice that types 14, 15, and 16 are land use categories, while the remainder are land cover categories. Approximately 50 pixels at various points within the image have no ground truth specified. The neural networks neither train nor test on those pixels.

As far as the network is concerned, the ground truth label for each pixel is assumed to be correct. Inaccuracies in the ground truth label with respect to a hypothetical true category hinder the ability of the network to learn the relevant features of each of the possible categories. For example, if some of the (true) water pixels were originally miscategorized as (ground truth) conifer trees, then the network will see two radically different profiles of (ground truth) conifer trees the first being those (true) conifer tree pixels correctly categorized as (ground truth) conifer and the second, those (true) water pixels miscategorized as (ground truth) conifer. The network might end up learning that conifer can look like (true) conifer or look like (true) water, in which case the network would tend to miscategorize all (true) water pixels as conifer. Even if the network was able to generalize all categories correctly, categorizing correctly even pixels whose ground truth label was incorrect, those latter pixels would still show up as incorrectly categorized in the overall performance statistics that are gathered. In the example just given, if the network correctly classified (real) water pixels as water, then the performance statistics would necessarily mark as mistakes those (real) water pixels incorrectly labeled as (ground truth) conifer.

A subset of the available data is used for training; the remainder of the data is used to test the network. Pixels which are bounded on all four sides with pixels having the same ground truth category as the center pixel are said to satisfy the non-boundary criterion (NB). Two training sets are defined: the first consisting of all pixels in the top half of the image which satisfy the NB criterion (termed TRAIN1); the second is all pixels in the top half of the image, regardless of the ground truth of their neighbors (termed TRAIN2). Therefore, the TRAIN1 set of pixels is a subset of the TRAIN2 set. The NB criterion eliminates from the training set those pixels forming the borders between categories. Because ground truth labeling must categorize pixels into only one type, border pixels are more likely to contain (in reality) multiple ground truths. In addition, this criteria will help compensate for errors in the ground truth file where pixels are shifted by one pixel or less. Of the 10,996 pixels in TRAIN2, only 4,405 pixels were accepted into training set TRAIN1 after application of this criteria. Therefore, 6,231 pixels in the top half of the image had at least one nearest neighbor in a different category. Two test sets are defined: the first consists of all pixels in the second half of the image which satisfy the same nearest neighbor criteria as the training set (termed TEST1); the second is all pixels in the second half of the image, regardless of the ground truths of their neighbors (termed TEST2). Therefore, the TEST1 set is a subset of the TEST2 set. There are 5,184 pixels in TEST1 and 10,997 pixels in TEST2 (see Table 1). By using two training sets and two test sets, we can determine whether the network has learned fundamental relationships between input data and ground truth, instead of just memorizing input-output pairs. Neural networks can learn any non-contradictory training set, given enough hidden nodes.

Table 1 lists the population of each type of ground truth in both training sets and both test sets. The image has no pixels of class 3, "standing corn". Because there are no pixels in the training sets of either class 1 or 15, "water" and "multiple family residential", respectively, pixels classed in these two categories will probably be miscategorized by a network. Since a neural network learns only those patterns on which it is trained, the classes in the training image should be representative of the whole image. An alternative of training upon pixels in the first half of the image is to train upon the first (or a random) half of the pixels in each category. Eventually, images of additional locations at different times of the year will also be used. In general, the power and robustness of a neural network system depends upon the breadth of its training sets.

There are four input nodes to the network, one for each spectral channel. Hidden nodes are completely connected to the input layer and to the output layer. The output layer uses one node for each of the 17 possible ground truths. After presentation of spectral values to the input layer, the output node, from 1 to 17, with the highest activation value indicates the network interpretation of the land use/cover category for the currently presented pixel. Called a unary encoding, one node for each ground truth type was chosen because adjacent ground truths (in the code) are not necessarily related types. If all types were mapped to one node, however, uncertainty of interpretation cannot be indicated by any activation values. For example, with unary encoding, activation values for nodes 1 and 15 of 0.5 might indicate uncertainty between type 1 and 15. If all types were mapped to one node, type 1 was mapped to activity value 0.1, and type 15 was mapped to activity value 0.9, then uncertainty between the two types would probably manifest itself with some intermediate value, signifying an unrelated type.

Table 1
Descriptions for ground truth codes and the population of these types in the TRAIN1 training set, the TRAIN2 training set, the TEST1 test set, and the TEST2 test set.

| | Number of Pixels in Set | | | | |
|---|---|---|---|---|---|
| Type | Train1 | Train2 | Test 1 | Test2 | Description of Ground Truth Type |
| 1 | 0 | 0 | 0 | 26 | water |
| 2 | 85 | 232 | 11 | 65 | agriculture - miscellaneous crops |
| 3 | 0 | 0 | 0 | 0 | corn - standing |
| 4 | 106 | 226 | 76 | 123 | corn - stubble |
| 5 | 74 | 206 | 45 | 312 | shrubland |
| 6 | 367 | 1307 | 670 | 1844 | grassland / pasture |
| 7 | 20 | 87 | 11 | 38 | soybeans |
| 8 | 19 | 77 | 126 | 426 | bare soil - cleared land |
| 9 | 2,492 | 4607 | 3,649 | 5,376 | hardwood forest, >70% of the forest component, 50-100% canopy closure |
| 10 | 138 | 589 | 65 | 317 | hardwood forest , >70% of the forest component, 10-50% canopy closure |
| 11 | 334 | 1238 | 224 | 830 | conifer forest, >50% of the forest component, 10-100% canopy closure |
| 12 | 62 | 362 | 24 | 237 | mixed wood forest, 10-100% canopy closure |
| 13 | 7 | 247 | 2 | 139 | asphalt |
| 14 | 524 | 1503 | 202 | 654 | residential - single family housing |
| 15 | 0 | 0 | 7 | 26 | residential - multiple family housing |
| 16 | 23 | 75 | 3 | 48 | industrial / commercial |
| 17 | 154 | 240 | 69 | 176 | bare soil - plowed field |
| Total | 4,405 | 10,996 | 5,184 | 10,997 | |

The input data are scaled linearly to a 0.1-0.9 range. A less than unit distance is chosen because activations of 0.0 or 1.0 require infinite weights. Infinite weights are undesirable because they require an infinite amount of time to learn. Scaling of the input is performed over a constant range for each channel: channel 1 is scaled between 65-165, channel 2 between 26-102, channel 3 between 20-127, and channel 4 between 55-151. These ranges were chosen so that, looking at a histogram of the data for a channel, all channel values with more than 1 pixel having that value would be inclusively within the scaled range. The result is saturation of some values with loss of potentially useful information. For example, the spectral data for the first pixel (ground truth 9) in training set TRAIN2 is 68, 26, 21, and 110 for channels 1 through 4 respectively. Those values would be scaled to 0.03, 0.0, 0.009, and 0.5 for channels 1 through 4 respectively for input to the network. If the spectral data for the second channel was 20 instead of 15, the scaled input for that channel would still have been 0.0. Using an input node for each possible spectral value for each channel would eliminate the need for any rescaling. This would multiply the number of input nodes by 256 from 4 to 1,024 and the number of connections accordingly. This is currently impractical because the large numbers of nodes and connections would require too much computer time.

The final performance of each network is measured by the proportion of the test pixels assigned to the correct land use category, the overall percentage correctly characterized (PCC). The PCC for each category type is also calculated. Anderson, et al. [AND76] suggest 85% as a minimum accuracy level of classification of remote sensor data. This level of performance is not expected at this stage.

Two investigations are described in this report. First, we determine a minimal number of hidden nodes sufficient to categorize the pixels in the training image and to categorize the test pixels as well as the training pixels. There are two reasons to minimize the number of hidden nodes: the first is that computation time for both training and testing increases with the number of hidden nodes; the second reason is that networks with too many hidden nodes can learn specifics of the training set which do not extend to the test set. The latter fault of networks with large numbers of hidden nodes is a consequence of too much representation power. Small number of hidden nodes "starve" a network into discovering the most parsimonious descriptions of regularities in the training set. Generally, simple generalizations extend to a test set more than complex generalizations. Networks with 1, 2, 3, 5, and 10 hidden nodes are trained on the TRAIN2 training set and tested on the TEST2 test set. Each network is trained 10,000 epochs with the momentum factor set to 0.5 and the network damping factor set to 0.5.

Next, we investigate the utility of the non-boundary criterion by training one 5 hidden node network on the TRAIN1 training set and another on the TRAIN2 training set. Each is trained for 10,000 epochs with the momentum factor set to 0.5 and the damping factor set to 0.5. Once training is complete, the networks are tested on both training sets and both test sets. This investigation determines the utility of limiting training to non-boundary pixels. Because those non-border points should be more separable than undifferentiated training pixels, it is expected that the PCC of non-border training pixels should be better than the PCC of undifferentiated training pixels. More importantly, because eliminating border points should reduce the amount of variance in the training set for each category, it is expected that training should be faster with non-border pixels than with undifferentiated pixels. In some cases, the network might not be able to extract the pattern in undifferentiated pixels, yet be able to do so if the training set is limited to non-boundary pixels. In that case, the PCC of the test set for networks trained on the non-boundary points would be superior to the PCC of the test set for networks trained on all pixels.

These experiments were performed on a SUN 4/280; every 1,000 epochs of training require 45, 54, 62, 78, and 118 seconds for networks with 1, 2, 3, 5, and 10 hidden nodes respectively. These times are not linearly proportional to the number of hidden nodes because of network overhead.

## Preliminary Results

Table 2 shows the percentage of correctly characterized pixels for each of the different land cover/use types for networks with a varying number of hidden nodes. Naturally, all the networks miscategorize those pixels in the test set with ground truth types that were not in the training set: types 1, 3, and 15. The overall PCC for the test set is better than the overall PCC for the training set for networks with 2 or greater hidden nodes. This is because the relative frequencies of the classes in the test set is different from that of the training set.

Table 2
True positive categorization percentages of the training set and test set for each ground truth type after training for networks with varying numbers of hidden nodes. Each is trained on the TRAIN2 training set for 10,000 epochs with the momentum factor set to 0.5 and the damping factor set to 0.5.

| | PCC - TRAIN2 | | | | | PCC - TEST2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Type | 1 | 2 | 3 | 5 | 10 | 1 | 2 | 3 | 5 | 10 |
| 1 | - | - | - | - | - | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
| 3 | - | - | - | - | - | - | - | - | - | - |
| 4 | 88 | 70 | 79 | 81 | 81 | 75 | 67 | 72 | 73 | 73 |
| 5 | 7 | 0 | 1 | 1 | 0 | 3 | 0 | 1 | 0 | 0 |
| 6 | 0 | 2 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 7 | 0 | 1 | 10 | 11 | 3 | 0 | 0 | 3 | 3 | 0 |
| 8 | 0 | 74 | 66 | 65 | 66 | 0 | 74 | 62 | 67 | 64 |
| 9 | 5 | 78 | 76 | 71 | 75 | 3 | 86 | 85 | 81 | 84 |
| 10 | 18 | 83 | 81 | 76 | 80 | 14 | 74 | 71 | 66 | 72 |
| 11 | 0 | 6 | 10 | 29 | 15 | 0 | 3 | 6 | 25 | 9 |
| 12 | 0 | 11 | 4 | 10 | 4 | 0 | 6 | 2 | 6 | 3 |
| 13 | 79 | 19 | 11 | 22 | 13 | 70 | 19 | 7 | 33 | 8 |
| 14 | 9 | 7 | 28 | 51 | 28 | 11 | 13 | 36 | 53 | 40 |
| 15 | - | - | - | - | - | 0 | 0 | 0 | 0 | 0 |
| 16 | 8 | 73 | 71 | 63 | 71 | 12 | 83 | 73 | 73 | 71 |
| 17 | 5 | 5 | 2 | 0 | 0 | 6 | 6 | 7 | 3 | 2 |
| Overall | 8 | 42 | 44 | 48 | 44 | 5 | 52 | 52 | 55 | 52 |

Because of the large amount of type 9 pixels, the PCC of that type has a large impact on the overall PCC. Whether a class has a high PCC or a low PCC is not dependent upon the relative frequency of those classes in the training set. This is due to the fact that the training method trains on the same number of pixels from each category during each epoch.

The 1 hidden node network can discriminate between classes 4 and 13 in the test set with PCCs of 75% and 70%, respectively. The PCCs for these classes are better than any network with more hidden nodes. Adding more hidden nodes does allow more classes to be distinguished, yet the ability to distinguish more classes comes at the expense of a decrease in the PCC of those already discriminated classes. There is no significant difference in the overall PCC beyond 1 hidden node. This suggests using many small 2 hidden node networks in unison, each network trained to discriminate between only a few categories.

Overall network performance for a network is best shown as a contingency table which shows the category into which each pixel has been placed, as a function of the correct ground truth for that pixel. A contingency table for the 5 hidden node network is shown in Table 3. The number in the $m^{th}$ row and $n^{th}$ column in the table indicates that percentage of pixels of ground truth $n$ which were categorized as class $m$. The true positive characterized percentages of each category type lie on the main diagonal. This table illustrates that pixels tend to be mischaracterized in a non-random fashion. For example, although the PCC for type

Table 3

Contingency table showing the percentage of pixels in each ground truth type categorized as type 1, 2, 3, ..., 17 in test set TEST2 for a 5 hidden node network after training for 10,000 epochs on training set TRAIN2. The columns, one for each ground truth, sum to 100%. There is one row for each network categorization possibility. For example, the entry 19 in row 4 and column 5 indicates that 19% of the ground truth type 5 pixels were mischaracterized as type 4. The true positive categorization percentages are along the main diagonal and shown in bold print.

| | Ground Truths | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Categorized | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
| 1 | **0** | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | **0** | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 34 | | **73** | 19 | 24 | 5 | 5 | 3 | 0 | 1 | 18 | 2 | 8 | 0 | 0 | 74 |
| 5 | 0 | 0 | | 0 | **0** | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | | 0 | 0 | **0** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 0 | 4 | | 1 | 1 | 7 | **3** | 5 | 0 | 0 | 0 | 5 | 3 | 6 | 0 | 0 | 3 |
| 8 | 39 | 3 | | 15 | 1 | 5 | 5 | **67** | 1 | 0 | 1 | 1 | 0 | 1 | 35 | 12 | 0 |
| 9 | 0 | 6 | | 0 | 60 | 13 | 34 | 1 | **81** | 20 | 27 | 41 | 0 | 2 | 0 | 0 | 17 |
| 10 | 14 | 10 | | 2 | 6 | 3 | 13 | 1 | 4 | **66** | 39 | 5 | 5 | 9 | 0 | 0 | 0 |
| 11 | 0 | 13 | | 3 | 7 | 8 | 24 | 0 | 8 | 11 | **25** | 18 | 1 | 4 | 0 | 0 | 0 |
| 12 | 0 | 6 | | 3 | 1 | 11 | 11 | 1 | 1 | 1 | 2 | **6** | 2 | 5 | 0 | 0 | 1 |
| 13 | 4 | 4 | | 0 | 1 | 5 | 0 | 2 | 0 | 1 | 1 | 0 | **33** | 7 | 0 | 0 | 1 |
| 14 | 4 | 17 | | 1 | 3 | 17 | 0 | 8 | 1 | 1 | 2 | 6 | 20 | **53** | 27 | 15 | 2 |
| 15 | 0 | 0 | | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | **0** | 0 | 0 |
| 16 | 39 | 1 | | 0 | 0 | 2 | 0 | 6 | 0 | 0 | 0 | 0 | 34 | 4 | 38 | **73** | 0 |
| 17 | 0 | 1 | | 2 | 1 | 3 | 5 | 3 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | **3** |

17, bare soil - plowed fields is only 3%, most of those pixels are classified as type 4 - corn stubble. Such miscategorization is reasonable, considering the close relationship between plowed fields and corn stubble. Recall that the process of photointerpretation to determine the ground truths did not consider variability within agricultural fields.

As examples of pictures of two network types, one categorized well and another not so well, Figure 1 shows pictures of the ground truths and the network categorizations for types 9 and 14, dense hardwood forest and single-family residential, respectively. The 5 hidden node network was used to categorize all the pixels in the image after training on the TRAIN2 training set. The general pattern of the residential area, type 14, is evident in the network categorization of that type (1*d*) despite the network having only a 53% PCC in the test region TEST2 for that type. Notice that a road like line extending from the top of the image to the lower left is categorized as residential. Not shown in this figure is that the network also categorizes other pixels along this same line as asphalt. Looking at the contingency table for this network (Table 3) one can see that 20% of the asphalt (type 13) pixels were miscategorized as residential (type 14). This is possibly explained by the network becoming confused by the close relationship between roads and residential areas in the apparent subdivision in the upper right of the image which is part of the training region.

The second investigation looked into the utility of limiting training to non-boundary pixels (see Table 4). Several observations about these results can be made. First, the PCC of the training pixels are higher for the network trained on the NB pixels, TRAIN1, than for the network trained on non-distinguished pixels, TRAIN2: overall PCC of 66% vs 48%, respectively. Therefore, during training it is easier to judge that the network training on the TRAIN1 set is learning the training set than it is to judge that the network training on the TRAIN2 set is learning the training set. Yet, when the network trained on the TRAIN2 set is subsequently tested on the TRAIN1 set, the overall PCC of 65% is comparable to the PCC for the network trained on the TRAIN1 training set of 66%.

The PCC on the tests sets are comparable between the two networks. The network trained on the TRAIN1 test set scores a 70% and 51 % overall PCC on test sets TEST1 and TEST2, respectively. The network trained on the TRAIN2 test set scores a 73% and 53% overall PCC on test sets TEST1 and TEST2, respectively. From this, it is evident that training on the TEST1 test set does not allow the network to discover patterns inherent in the data any better than training on the TEST2 test set. Yet, there is an obvious difference between the networks ability to classify non-border pixels and non-distinguished pixels. Apparently, the network trained on the TRAIN2 training set is able to filter out the noise endemic to the border pixels, or at least that noise is overshadowed by the signal in the non-border pixels.

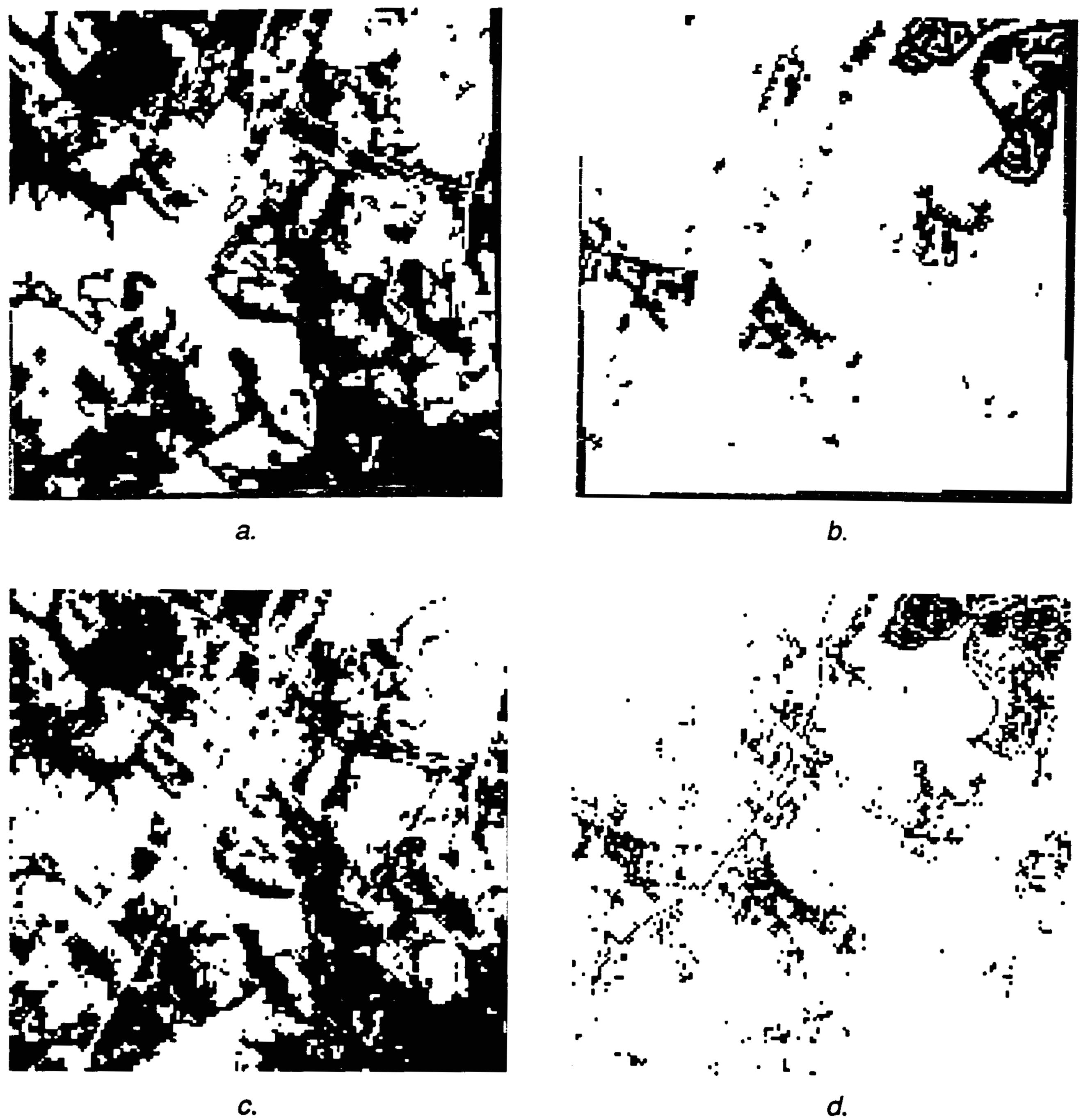


Fig. 1. A pictorial comparison of the categorization performance of the 5 hidden node network for category types 9 and 14, dense hardwood forest and residential respectively. Network was trained on top half of image. Sub-figures *a* and *b* show the pixels with ground truth types 9 and 14. Sub-figures *c* and *d* show the pixels that the 5 hidden node neural network classified as types 9 and 14. Pixels positive for the specified type are shown black (small white dots on black are meaningless), all other pixels are shown clear. The irregular border of the ground truth is shown by a lighter black background (*a* and *b*). That border area is filled in by the network because spectral information is available for a larger area (*c* and *d*).

Table 4
Percent correct characterized (PCC) of the training sets and the test sets after training with the training sets TRAIN1 and TRAIN2. Each of the 5 hidden node networks is trained for 10,000 epochs with the momentum factor set to 0.5 and the damping factor set to 0.5.

| | Network Trained on TRAIN1 | | | | Network Trained on TRAIN2 | | | |
|---|---|---|---|---|---|---|---|---|
| Type | TRAIN1 | TRAIN2 | TEST1 | TEST2 | TRAIN1 | TRAIN2 | TEST1 | TEST2 |
| 1 | - | - | - | 0 | - | - | - | 0 |
| 2 | 4 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | - | - | - | - | - | - | - | - |
| 4 | 96 | 78 | 87 | 70 | 96 | 81 | 91 | 73 |
| 5 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | 45 | 29 | 0 | 0 | 10 | 11 | 0 | 3 |
| 8 | 100 | 65 | 81 | 58 | 100 | 65 | 87 | 67 |
| 9 | 80 | 61 | 86 | 74 | 88 | 71 | 92 | 81 |
| 10 | 95 | 70 | 82 | 57 | 99 | 76 | 82 | 66 |
| 11 | 59 | 49 | 50 | 45 | 29 | 29 | 28 | 25 |
| 12 | 27 | 9 | 12 | 8 | 2 | 10 | 4 | 6 |
| 13 | 71 | 35 | 100 | 43 | 0 | 22 | 0 | 33 |
| 14 | 79 | 76 | 86 | 75 | 52 | 51 | 65 | 53 |
| 15 | - | - | 0 | 0 | - | - | 0 | 0 |
| 16 | 0 | 0 | 0 | 0 | 70 | 63 | 67 | 73 |
| 17 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 3 |
| Overall | 66 | 49 | 70 | 51 | 65 | 48 | 73 | 53 |

**Preliminary Conclusions**

From these results it can be concluded that information is available in the raw spectral values concerning the ground truth classes into which individual pixels can be categorized. The limited training set precludes any conclusions about how accurate a neural network could get with this data. With some qualifications, a comparison can be made between the results of this study and and results of the earlier study for which the ground truths were originally collected [WIL84]. Although that earlier study used six bands of Thematic Mapper imagery compared to only four for this study, that study used imagery taken on November 2, a time "far from optimal for general category discrimination" [WIL84]. The current study used only 4 bands because at the time the image was taken, in July, 1982, the other instruments were not yet functioning. Because the earlier study used pixels from 9 sub-regions, only the first sub-region being used in this study, that earlier study had more training pixels than this study: 1600 training pixels and 600 test pixels for each class were chosen (see Table 1 for the number of training pixels in this study). When that early study used an iterative, point migration clustering algorithm with no editing of the training statistics, the overall PCC was 36.7%. When the analyst was allowed to interact with the computer to edit training statistics, then classification accuracy rose to 62.0%. When the 17 categories were aggregated into five categories (water, crops, pasture and grass, forest, and urban) and no editing of the training statistics was used, then classification accuracy was 65.7%. Because the neural network does not require any user interaction during training it is appropriate to conclude that this approach, with an overall PCC of 52.1% for the 2 hidden node network after 10,000 epochs, is quite favorable, despite the fact that the LANDSAT images were taken at different times for the two studies.

## Future Directions

Because of our eventual goal to use this system for real time data characterization and categorization, input data should undergo minimal preprocessing. Among other uses, preprocessing can change the data into forms amenable for neural network processing, smooth out extremes in spectral flux values, and provide derivative measures of the data such as texture. That first alternative, changing the form of the data by representation, is frequently done by transforming the data from the spatial domain to the frequency domain. Derivative measures could use image segmentation. In this regard, examinations of a segmented version of the image will be made. Tilton [TIL88] has developed a parallel region growing segmentation method which has been used to segment the image into 207 subregions. That process substituted the spectral data for each pixel with the average spectral data for all pixels in that pixel's sub-region. Thus, the average value of the spectral bands for pixels in the sub-regions will be used as input to the neural network. Other derivative measures of these regions, such as texture, can also be tested for their usefulness.

Another possible direction is to perform a fine discrimination by a sequence of course discriminations. If a network can consistently discriminate between groups of classes (superclasses) or individual classes, a hierarchy of such neural networks could be used to extract and refine discriminations between ground truths. For example, a network could categorize 5 mixed classes into 3 superclasses, one superclass consisting of one class, the other two superclasses consisting of 2 classes each. Successive networks could further refine each of the two superclasses into their respective class constituents. This is particularly suggested by the ability of the 1 hidden node network to successfully distinguish types 4 and 13 (see Table 2). The decision of when, and how, to apply different kinds of networks could eventually be implemented by an expert system.

In order to extend this paradigm to work on multiple images taken of different locations and from different instruments, the use of absolute spectral data must be modified. The network will have to use relative shape of the spectral flux curve. One possible way to implement this is by a form of lateral inhibition among the input layer nodes.

## Bibliography

[AND76] J. R. Anderson, E. E. Hardy, J. T. Roach, and R. E. Witmer, "A land use and land cover classification system for use with remote sensor data," *Geological Survey Professional Paper 964,* United States Government Printing Office: Washington, 1976.

[CAM88] W. J. Campbell, L. Roelofs, and M. Goldberg, "Automated cataloging and characterization of space-derived data," *Telematics and Informatics,* Vol. 5 No. 3, pp 279-288, 1988.

[FOG88] R. J. Fogler, R. L. Williams, and L. D. Hostetler, "A billion connection per second feedforward pipeline for computer vision applications," *Abstracts of the First Annual INNS Meeting,* Boston, Sept. 6-10,1988.

[RUM85] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," In D. E. Rumelhart, J.L. McClelland and the PDP research group (editors), *Parallel distributed processing: Explorations in the microstructure of cognition. Volume I: Foundations.* Cambridge, MA: MIT Press, 1986.

[TIL88] J. C. Tilton, "Image segmentation by iterative parallel region growing with applications to data compression and image analysis," *Proceeding of Second Symposium on Frontiers of Massively Parallel Computation,* Fairfax VA, Oct. 10-12, 1988.

[TIL89] J. C. Tilton, personal communication, 1989.

[WIL84] D. L. Williams, J. R. Irons, B. L. Markham, R. F. Nelson, D. L. Toll, R. S. Latty, and M. L. Stauffer, "A statistical evaluation of the advantages of LANDSAT thematic mapper data in comparison to multispectral scanner data," *IEEE Trans. on Geoscience and Remote Sensing,* GE-22, pp. 294-302, 1984.